Available online at www.sciencedirect.com
ScienceDirect
ELSEVIER
Journal of Approximation Theory 158 (2009) 170–183
Journal of Approximation Theory
www.elsevier.com/locate/jat

Bregman distances and Klee sets

Heinz H. Bauschke[a,*], Xianfu Wang[a], Jane Ye[b], Xiaoming Yuan[c]

[a] *Mathematics, Irving K. Barber School, The University of British Columbia Okanagan, Kelowna, B.C. V1V 1V7, Canada*
[b] *Department of Mathematics and Statistics, University of Victoria, Victoria, B.C. V8P 5C2, Canada*
[c] *Department of Mathematics, Hong Kong Baptist University, PR China*

Received 16 February 2008; received in revised form 24 July 2008; accepted 1 August 2008
Available online 12 November 2008

Communicated by C.K. Chui and H.N. Mhaskar

Dedicated to the memory of G.G. Lorentz

Abstract

In 1960, Klee showed that a subset of a Euclidean space must be a singleton provided that each point in the space has a unique farthest point in the set. This classical result has received much attention; in fact, the Hilbert space version is a famous open problem. In this paper, we consider Klee sets from a new perspective. Rather than measuring distance induced by a norm, we focus on the case when distance is meant in the sense of Bregman, i.e., induced by a convex function. When the convex function has sufficiently nice properties, then – analogously to the Euclidean distance case – every Klee set must be a singleton. We provide two proofs of this result, based on Monotone Operator Theory and on Nonsmooth Analysis. The latter approach leads to results that complement the work by Hiriart-Urruty on the Euclidean case.

Keywords: Convex function; Legendre function; Bregman distance; Bregman projection; Farthest point; Maximal monotone operator; Subdifferential operator

1. Introduction

Throughout this paper, $\mathbb{R}^J$ denotes the standard Euclidean space with inner product $\langle \cdot, \cdot \rangle$ and induced norm $\| \cdot \|$. Let C be a nonempty bounded closed subset of $\mathbb{R}^J$ and assume that C is

* Corresponding author.
E-mail addresses: heinz.bauschke@ubc.ca (H.H. Bauschke), shawn.wang@ubc.ca (X. Wang), janeye@math.uvic.ca (J. Ye), xmyuan@hotmail.com (X. Yuan).

0021-9045/$ - see front matter
doi:10.1016/j.jat.2008.08.015

a *Klee set (with respect to the Euclidean distance)*, i.e., each point in $\mathbb{R}^J$ has a unique farthest point in C. Must C be a singleton? The *farthest-point conjecture* [11] states that the answer to this question is affirmative. This conjecture has attracted many mathematicians; see, e.g., [4,10–13,25] and the references therein. Although the farthest-point conjecture is true in $\mathbb{R}^J$, as was shown originally by Klee [14] (see also [1,11,17]), only partial results are known in infinite-dimensional settings (see, e.g., [18,25]). While the farthest-point conjecture is primarily of theoretical interest, it should be noted that farthest points do play a role in computational geometry; see, e.g., the sections on Voronoi diagrams in [20].

In this paper, we cast a new light on this problem by measuring the distance in the sense of Bregman rather than in the usual Euclidean sense. To this end, assume that

$$f : \mathbb{R}^J \to \,]-\infty, +\infty] \text{ is convex and differentiable on } U := \operatorname{int}\operatorname{dom} f \neq \varnothing, \tag{1}$$

where $\operatorname{int}\operatorname{dom} f$ stands for the interior of the set $\operatorname{dom} f := \{x \in \mathbb{R}^J \mid f(x) \in \mathbb{R}\}$. Then the *Bregman distance* [5] with respect to f, written as D_f or simply D, is

$$D : \mathbb{R}^J \times \mathbb{R}^J \to [0, +\infty] : (x, y) \mapsto \begin{cases} f(x) - f(y) - \langle \nabla f(y), x - y \rangle, & \text{if } y \in U; \\ +\infty, & \text{otherwise.} \end{cases} \tag{2}$$

Although standard, it is well known that the name "Bregman distance" is somewhat misleading because in general D is neither symmetric nor does the triangle inequality hold. We recommend books [6,7] to the reader for further information on Bregman distances and their various applications.

Throughout, we assume that

$$C \subset U. \tag{3}$$

Now define the *left-Bregman-farthest-distance function* by

$$\overleftarrow{F}_C : U \to [0, +\infty] : y \mapsto \sup_{x \in C} D(x, y), \tag{4}$$

and the corresponding *left-Bregman-farthest-point map* by

$$\overleftarrow{Q}_C : U \to U : y \mapsto \operatorname*{argmax}_{x \in C} \; D(x, y). \tag{5}$$

Since D is in general not symmetric, there exist analogously the *right-Bregman-farthest-distance function* and the *right-Bregman-farthest-point map*. These objects, which we will study later, are denoted by $\overrightarrow{F}_C$ and $\overrightarrow{Q}_C$, respectively. When $f = \frac{1}{2}\|\cdot\|^2$, then $D : (x, y) \mapsto \frac{1}{2}\|x - y\|^2$ is symmetric and the corresponding map $\overleftarrow{Q}_C$ is identical to the farthest-point map with respect to the Euclidean distance.

The present more general framework based on Bregman distances allows for significant extensions of Hiriart-Urruty's work [11] (and for variants of some of the results in [25]). One of our main results states that if f is sufficiently nice, then every Klee set (with respect to D) must be a singleton. Two fairly distinct proofs of this result are given. The first is based on the deep Brézis–Haraux range approximation theorem from monotone operator theory. The second proof, which uses generalized subdifferentials from nonsmooth analysis, allows us to characterize sets with unique farthest points. Various subdifferentiability properties of the Bregman-farthest-distance function are also provided. The present work complements a corresponding study on Chebyshev sets [3], where the focus is on *nearest* rather than farthest points.

The paper is organized as follows. In Section 2, we state our standing assumptions and we provide some concrete examples for f. In Section 3, Bregman farthest points are characterized and it is shown that the Bregman-farthest-distance function is locally Lipschitz. The first proof of our main result is presented in Section 4. In Section 5, we study subdifferentiabilities of the farthest-distance function. We establish Clarke regularity, and we provide an explicit formula for the Clarke subdifferential. Section 6 contains several characterizations of Klee sets. The results extend Hiriart-Urruty's work [11] from Euclidean to Bregman distances. In the final Section 7, we show that the right-Bregman-farthest-point map $\overrightarrow{Q}_C^f$ can be studied in terms of the left and dual counterpart $\overleftarrow{Q}_{\nabla f(C)}^{f^*}$. When f is sufficiently nice, this allows us to deduce that Klee sets with respect to the right-Bregman-farthest-point map are necessarily singletons.

We employ standard notation from Convex Analysis; see, e.g., [21,22,26]. For a function h, the subdifferential in the convex-analytical sense is denoted by ∂h, h^* stands for the Fenchel conjugate, and $\operatorname{dom} h$ is the set of all points where h is not $+\infty$. If h is differentiable at x, then $\nabla h(x)$ and $\nabla^2 h(x)$ denote the gradient vector and the Hessian matrix at x, respectively. The notation $\operatorname{conv} h$ ($\overline{\operatorname{conv}} h$) denotes the convex hull (closed convex hull) of h. For a set S, the expressions $\operatorname{int} S$, $\operatorname{cl} S$, $\operatorname{conv} S$, $\overline{\operatorname{conv}} S$ signify the interior, the closure, the convex hull, and the closed convex hull of S, respectively. A set-valued operator T from X to Y, is written as $T\colon X \rightrightarrows Y$, and $\operatorname{dom} T$ and $\operatorname{ran} T$ stand for the domain and range of T. Finally, we simply write $\underline{\lim}$ and $\overline{\lim}$, for the limit inferior and limit superior (as they occur in set-valued analysis).

2. Standing assumptions and examples

From now on, and until the end of Section 6, our standing assumptions are:

A1 The function $f\colon \mathbb{R}^J \to]-\infty, +\infty]$ is a *convex function of Legendre type*, i.e., f is essentially smooth and essentially strictly convex in the sense of Rockafellar [21, Section 26], with $U := \operatorname{int}\operatorname{dom} f$.

A2 The function f is *1-coercive* (also known as *supercoercive*), i.e., $\lim_{\|x\|\to+\infty} f(x)/\|x\| = +\infty$. An equivalent requirement is $\operatorname{dom} f^* = \mathbb{R}^J$ (see [22, Theorem 11.8(d)]).

A3 The set C is a nonempty bounded closed (hence compact) subset of U.

There are many instances of functions satisfying **A1**–**A3**. We list only a few.

Example 2.1. Let $x = (x_j)_{1\le j\le J}$ and $y = (y_j)_{1\le j\le J}$ be two points in $\mathbb{R}^J$.

(i) *(Halved) Energy:* If $f = \frac{1}{2}\|\cdot\|^2$, then $U = \mathbb{R}^J$, $f^* = f$, and

$$D(x, y) = \frac{1}{2}\|x - y\|^2.$$

Thus, the Bregman distance with respect to the (halved) energy corresponds to the usual Euclidean distance.

(ii) *(Negative) Boltzmann–Shannon Entropy:* If $f\colon x \mapsto \sum_{j=1}^J x_j \ln(x_j) - x_j$ for $x \ge 0$, $+\infty$ otherwise (where $x \ge 0$ and $x > 0$ are understood coordinate-wise and $0 \ln 0 := 0$), then $U = \{x \in \mathbb{R}^J \mid x > 0\}$, $f^*(y) = \sum_{j=1}^J \exp(y_j)$, and

$$D(x, y) = \begin{cases} \displaystyle\sum_{j=1}^J x_j \ln(x_j/y_j) - x_j + y_j, & \text{if } x \ge 0 \text{ and } y > 0;\\ +\infty, & \text{otherwise} \end{cases}$$

is the famous *Kullback–Leibler Divergence*.

(iii) More generally, given a function $\phi\colon \mathbb{R} \to]-\infty, +\infty]$ satisfying **A1**, **A2** and setting $f(x) = \sum_{j=1}^{J} \phi(x_j)$, one has the same properties for f, with $U = (\operatorname{int}\operatorname{dom}\phi)^J$ and

$$D(x, y) = \sum_{j=1}^{J} \phi(x_j) - \phi(y_j) - \phi'(y_j)(x_j - y_j).$$

For instance, one may consider $\phi\colon t \mapsto |t|^p/p$, where $p > 1$.

The following result recalls a key property of Legendre functions.

Fact 2.2 (*Rockafellar [21, Theorem 26.5]*). If h is a convex function of Legendre type, then so is h^* and

$$\nabla h : \operatorname{int}\operatorname{dom} h \to \operatorname{int}\operatorname{dom} h^*$$

is a topological isomorphism with inverse mapping $(\nabla h)^{-1} = \nabla h^*$.

Corollary 2.3. *The mappings $\nabla f\colon U \to \mathbb{R}^J$ and $\nabla f^*\colon \mathbb{R}^J \to U$ are continuous, bijective, and inverses of each other.*

3. Left Bregman farthest distances and farthest points

The following result generalizes Hiriart-Urruty's [11, Proposition 3.1 and Corollary 3.2] and provides a characterization of left-Bregman-farthest-points (recall (4) and (5)).

Proposition 3.1. *Let $y \in U$, $x \in C$, and $\lambda \geq 1$. Then*

$$x \in \overleftarrow{Q}_C(y) \Leftrightarrow (\forall c \in C) D(c, x) \leq \langle \nabla f(y) - \nabla f(x), c - x\rangle. \tag{6}$$

If $x \in \overleftarrow{Q}_C(y)$ and

$$z_\lambda := \nabla f^*(\lambda \nabla f(y) + (1 - \lambda)\nabla f(x)), \tag{7}$$

then $x \in \overleftarrow{Q}_C(z_\lambda)$; moreover, if $\lambda > 1$, then $\overleftarrow{Q}_C(z_\lambda) = \{x\}$.

Proof. By definition, $x \in \overleftarrow{Q}_C(y)$ means that for each $c \in C$, $0 \geq D(c, y) - D(x, y)$, i.e.,

$$\begin{aligned} 0 &\geq f(c) - f(x) - \langle \nabla f(y), c - x\rangle \\ &= f(c) - f(x) - \langle \nabla f(x), c - x\rangle + \langle \nabla f(x) - \nabla f(y), c - x\rangle \\ &= D(c, x) - \langle \nabla f(y) - \nabla f(x), c - x\rangle. \end{aligned}$$

Hence (6) follows. Now assume that $x \in \overleftarrow{Q}_C(y)$ and take an arbitrary $c \in C$. By (6),

$$\langle \nabla f(y) - \nabla f(x), c - x\rangle \geq 0. \tag{8}$$

The definition of z_λ and (8) result in

$$\langle \nabla f(z_\lambda) - \nabla f(x), c - x\rangle = \lambda\langle \nabla f(y) - \nabla f(x), c - x\rangle \geq \langle \nabla f(y) - \nabla f(x), c - x\rangle. \tag{9}$$

Now (6) and (9) imply

$$D(c, x) \leq \langle \nabla f(y) - \nabla f(x), c - x\rangle \leq \langle \nabla f(z_\lambda) - \nabla f(x), c - x\rangle, \tag{10}$$

which – again by (6) – yields that $x \in \overleftarrow{Q}_C(z_\lambda)$. Finally, assume that $\lambda > 1$ and let $\hat{x} \in \overleftarrow{Q}_C(z_\lambda)$. By (7), $x \in \overleftarrow{Q}_C(z_\lambda)$. Since $D(x, z_\lambda) = D(\hat{x}, z_\lambda)$, we have

$$\begin{aligned} 0 &= D(x, z_\lambda) - D(\hat{x}, z_\lambda) \\ &= f(x) - f(\hat{x}) - \langle \nabla f(z_\lambda), x - \hat{x} \rangle \\ &= f(x) - f(\hat{x}) - \langle \lambda \nabla f(y) + (1-\lambda) \nabla f(x), x - \hat{x} \rangle \end{aligned}$$

so that

$$(1-\lambda)[f(x) - f(\hat{x}) - \langle \nabla f(x), x - \hat{x} \rangle] + \lambda [f(x) - f(\hat{x}) - \langle \nabla f(y), x - \hat{x} \rangle] = 0.$$

Then $(1-\lambda)[f(\hat{x}) - f(x) - \langle \nabla f(x), \hat{x} - x \rangle] = \lambda [f(x) - f(\hat{x}) - \langle \nabla f(y), x - \hat{x} \rangle]$, and thus

$$(1-\lambda) D(\hat{x}, x) = \lambda [D(x, y) - D(\hat{x}, y)].$$

It follows that

$$D(x, y) - D(\hat{x}, y) = \frac{1-\lambda}{\lambda} D(\hat{x}, x). \tag{11}$$

Assume that $x \neq \hat{x}$. Then $D(\hat{x}, x) > 0$, and, since $\lambda > 1$, we get $0 > (1-\lambda) D(\hat{x}, x)$. In view of (11), we conclude that $D(x, y) < D(\hat{x}, y)$, which contradicts that x is a farthest point of y. Therefore, $x = \hat{x}$. ■

It will be convenient to define $f^\vee = f \circ (-\mathrm{Id})$, i.e., $f^\vee(y) = f(-y)$ for every $y \in \mathbb{R}^J$. Our standing assumptions **A1**–**A3** imply that the function

$$-f^\vee + \iota_{-C} \colon \mathbb{R}^J \to \left]-\infty, +\infty\right] \colon x \mapsto \begin{cases} -f(-x), & \text{if } x \in -C; \\ +\infty, & \text{otherwise} \end{cases} \tag{12}$$

is lower semicontinuous. This function plays a role in our next result, where we show that $\overleftarrow{F}_C$ is a locally Lipschitz function on U.

Proposition 3.2. *The left-Bregman-farthest-distance function $\overleftarrow{F}_C$ is continuous on U and it can be written as the composition*

$$\overleftarrow{F}_C = \left(f^* + (-f^\vee + \iota_{-C})^*\right) \circ \nabla f, \tag{13}$$

where $f^* + (-f^\vee + \iota_{-C})^*$ is locally Lipschitz and ∇f is continuous. Consequently, $\overleftarrow{F}_C$ is locally Lipschitz on U provided that ∇f has the same property—as is the case when f is twice continuously differentiable. Finally,

$$(-f^\vee + \iota_{-C})^* = \overleftarrow{F}_C \circ \nabla f^* - f^*, \tag{14}$$

and hence $\overleftarrow{F}_C \circ \nabla f^*$ is a locally Lipschitz convex function with full domain.

Proof. Fix $y \in U$. Then

$$\begin{aligned}
\overleftarrow{F}_C(y) &= \sup_{c\in C}[f(c) - f(y) - \langle \nabla f(y), c - y\rangle] \\
&= \sup_{c\in C}[f(c) - \langle \nabla f(y), c\rangle] + f^*(\nabla f(y)) \\
&= f^*(\nabla f(y)) + \sup_{c\in C}[\langle \nabla f(y), -c\rangle - (-f)(c)] \\
&= f^*(\nabla f(y)) + \sup_{c}[\langle \nabla f(y), -c\rangle - (-f(c) + \iota_C(c))] \\
&= f^*(\nabla f(y)) + \sup_{z}[\langle \nabla f(y), z\rangle - (-f(-z) + \iota_{-C}(z))] \\
&= f^*(\nabla f(y)) + (-f^{\vee} + \iota_{-C})^*(\nabla f(y)).
\end{aligned}$$

Assumptions **A1–A3** imply that $-f^{\vee} + \iota_{-C}$ is proper and 1-coercive. By [13, Proposition X.1.3.8], the convex function $(-f^{\vee} + \iota_C)^*$ has full domain and it thus is locally Lipschitz on $\mathbb{R}^J$. Since f^* is likewise locally Lipschitz on $\mathbb{R}^J$, Fact 2.2 yields the continuity of $\overleftarrow{F}_C$. The "Consequently" statement is a consequence of the Mean Value Theorem. Finally, precomposing (13) by ∇f^* followed by rearranging yields (14), which in turn shows that $\overleftarrow{F}_C \circ \nabla f^*$ is a locally Lipschitz convex function, as it is the sum of two such functions. ■

4. Left-Bregman-farthest-point maps

The next result contains some useful properties of the farthest-point map and item (iii) is an extension of [11, Proposition 3.3].

Proposition 4.1. *Let x and y be in U. Then the following hold.*

(i) $\overleftarrow{Q}_C(x) \neq \varnothing$.

(ii) *If $(x_n)_{n\in\mathbb{N}}$ is a sequence in U converging to x and $(c_n)_{n\in\mathbb{N}}$ is a sequence in C such that $(\forall n \in \mathbb{N})$ $c_n \in \overleftarrow{Q}_C(x_n)$, then all cluster points of $(c_n)_{n\in\mathbb{N}}$ lie in $\overleftarrow{Q}_C(x)$. Consequently, $\overleftarrow{Q}_C : U \rightrightarrows C$ is compact-valued and upper semicontinuous (in the sense of set-valued analysis).*

(iii) $\langle -\overleftarrow{Q}_C(x) + \overleftarrow{Q}_C(y), \nabla f(x) - \nabla f(y)\rangle \geq 0$ *and hence* $-\overleftarrow{Q}_C \circ \nabla f^*$ *is monotone.*

Proof. (i) Since $D(\cdot, x)$ is continuous on U and C is a compact subset of U, it follows that $D(\cdot, x)$ attains its supremum over C.

(ii) Suppose that $(x_n)_{n\in\mathbb{N}}$ lies in U and converges to x, that $(c_n)_{n\in\mathbb{N}}$ lies in C, and that $(\forall n \in \mathbb{N})$ $c_n \in \overleftarrow{Q}_C(x_n)$, i.e.,

$$(\forall n \in \mathbb{N}) \quad f(c_n) - f(x_n) - \langle \nabla f(x_n), c_n - x_n\rangle = D(c_n, x_n) = \overleftarrow{F}_C(x_n). \tag{15}$$

By **A3**, $(c_n)_{n\in\mathbb{N}}$ has cluster points and they all lie in C. After passing to a subsequence if necessary, we assume that $c_n \to \bar{c} \in C$. Since $\overleftarrow{F}_C$ is continuous on U by Proposition 3.2, we pass to the limit in (15) and deduce that $f(\bar{c}) - f(x) - \langle \nabla f(x), \bar{c} - x\rangle = D(\bar{c}, x) = \overleftarrow{F}_C(x)$. Hence $\bar{c} \in \overleftarrow{Q}_C(x)$. The same reasoning (with $(x_n)_{n\in\mathbb{N}} = (x)_{n\in\mathbb{N}}$) shows that $\overleftarrow{Q}_C(x)$ is closed and hence compact (since C is compact). Therefore, $\overleftarrow{Q}_C$ is compact-valued and upper semicontinuous on U.

(iii) Let $p \in \overleftarrow{Q}_C x$ and $q \in \overleftarrow{Q}_C y$. Then $D(p, x) \geq D(q, x)$ and $D(q, y) \geq D(p, y)$. Using (2), we obtain $f(p) - f(q) - \langle \nabla f(x), p - q \rangle \geq 0$ and $f(q) - f(p) - \langle \nabla f(y), q - p \rangle \geq 0$. Adding these two inequalities yields $\langle \nabla f(x) - \nabla f(y), q - p \rangle \geq 0$. The result now follows from Corollary 2.3. ■

Definition 4.2. The set C is *Klee with respect to the left Bregman distance*, or simply $\overleftarrow{D}$*-Klee*, if for every $x \in U$, $\overleftarrow{Q}_C(x)$ is a singleton.

Proposition 4.3. *Suppose that C is $\overleftarrow{D}$-Klee. Then $\overleftarrow{Q}_C : U \to C$ is continuous. Hence $-\overleftarrow{Q}_C \circ \nabla f^*$ is continuous and maximal monotone.*

Proof. By Proposition 4.1(ii), $\overleftarrow{Q}_C$ is continuous on U. This and the continuity of $\nabla f^* : \mathbb{R}^J \to U$ (see Corollary 2.3) imply that $-\overleftarrow{Q}_C \circ \nabla f^*$ is continuous. On the other hand, Proposition 4.1(iii) shows that $-\overleftarrow{Q}_C \circ \nabla f^*$ is monotone. Altogether, using [22, Example 12.7], we conclude that $-\overleftarrow{Q}_C \circ \nabla f^*$ is maximal monotone on $\mathbb{R}^J$. ■

The Brézis–Haraux range approximation theorem plays a crucial role in the proof of the following main result. It is interesting to note that the Hilbert space analogue [25, Proposition 6.2] by Westphal and Schwartz relies only on the less powerful Minty's theorem.

Theorem 4.4 ($\overleftarrow{D}$*-Klee-Sets are Singletons*). *Suppose that C is $\overleftarrow{D}$-Klee. Then C is a singleton.*

Proof. Recall Corollary 2.3 and consider the following two maximal monotone operators (see Proposition 4.3) ∇f^* and $-\overleftarrow{Q}_C \circ \nabla f^*$. The Brézis–Haraux range approximation theorem (see [23, Section 19]) implies that

$$\begin{aligned} \operatorname{int} \operatorname{ran} \Big(\nabla f^* - (\overleftarrow{Q}_C \circ \nabla f^*) \Big) &= \operatorname{int} \Big(\operatorname{ran} \nabla f^* - \operatorname{ran}(\overleftarrow{Q}_C \circ \nabla f^*) \Big) \\ &= \operatorname{int} \Big(U - \operatorname{ran}(\overleftarrow{Q}_C \circ \nabla f^*) \Big). \end{aligned} \tag{16}$$

Since $\operatorname{ran}(\overleftarrow{Q}_C \circ \nabla f^*) \subseteq C$ and $C \subset U$, we have $0 \in \operatorname{int}(U - \operatorname{ran}(\overleftarrow{Q}_C \circ \nabla f^*))$, and hence, by (16), $0 \in \operatorname{int} \operatorname{ran}(\nabla f^* - (\overleftarrow{Q}_C \circ \nabla f^*))$. Thus there exists $x \in \mathbb{R}^J$ such that $\overleftarrow{Q}_C(\nabla f^*(x)) = \nabla f^*(x)$. Hence C must be a singleton. ■

Corollary 4.5. *The set C is $\overleftarrow{D}$-Klee if and only if it is a singleton.*

5. Subdifferentiability properties

For a function g that is finite and locally Lipschitz at a point $y \in \mathbb{R}^J$, the *Dini subderivative* and the *Clarke subderivative* of g at y in the direction $w \in \mathbb{R}^J$, denoted respectively by $\mathrm{d}g(y)(w)$ and $\hat{\mathrm{d}}g(y)(w)$, are defined via

$$\mathrm{d}g(y)(w) := \varliminf_{t \downarrow 0} \frac{g(y + tw) - g(y)}{t},$$

$$\hat{\mathrm{d}}g(y)(w) := \varlimsup_{\substack{x \to y \\ t \downarrow 0}} \frac{g(x + tw) - g(x)}{t},$$

and the corresponding *Dini subdifferential* and *Clarke subdifferential* via

$$\hat{\partial} g(y) := \{y^* \in \mathbb{R}^J \mid (\forall w \in \mathbb{R}^J) \langle y^*, w\rangle \leq \mathrm{d}g(y)(w)\},$$
$$\overline{\partial} g(y) := \{y^* \in \mathbb{R}^J \mid (\forall w \in \mathbb{R}^J) \langle y^*, w\rangle \leq \hat{\mathrm{d}}g(y)(w)\}.$$

The *limiting subdifferential* (see [22, Definition 8.3]) is defined by

$$\partial_L g(y) := \varlimsup_{x \to y} \hat{\partial} g(x).$$

We say that g is *Clarke regular* at y if $\mathrm{d}g(y)(w) = \hat{\mathrm{d}}g(y)(w)$ for every $w \in \mathbb{R}^J$, or equivalently $\hat{\partial} g(y) = \overline{\partial} g(y)$. For further properties of these subdifferentials and subderivatives, see [8,16,22].

We now provide various subdifferentiability properties of $\overleftarrow{F}_C$ in terms of $\overleftarrow{Q}_C$, and show that $\overleftarrow{F}_C$ is Clarke regular.

Proposition 5.1 (*Clarke Regularity*). *Suppose that f is twice continuously differentiable on U, and let $y \in U$. Then*

$$(\forall w \in \mathbb{R}^J) \quad \mathrm{d}\overleftarrow{F}_C(y)(w) = \hat{\mathrm{d}}\overleftarrow{F}_C(y)(w) = \max\langle \nabla^2 f(y)(y - \overleftarrow{Q}_C(y)), w\rangle \tag{17}$$

and

$$\partial_L \overleftarrow{F}_C(y) = \hat{\partial}\overleftarrow{F}_C(y) = \overline{\partial}\overleftarrow{F}_C(y) = \nabla^2 f(y)[y - \operatorname{conv} \overleftarrow{Q}_C(y)]; \tag{18}$$

consequently, $\overleftarrow{F}_C$ is Clarke regular on U.

Proof. Set $g := \overleftarrow{F}_C$ and let $x \in \overleftarrow{Q}_C(y)$. Fix $w \in \mathbb{R}^J$ and choose $t > 0$ sufficiently small so that $y + tw \in U$. Since $x \in \overleftarrow{Q}_C(y)$, we note that

$$\begin{aligned} g(y+tw) &\geq f(x) - f(y+tw) - \langle \nabla f(y+tw), x - (y+tw)\rangle \\ &= f(x) - f(y+tw) - \langle \nabla f(y+tw), x - y\rangle + \langle \nabla f(y+tw), tw\rangle \end{aligned}$$

and $g(y) = f(x) - f(y) - \langle \nabla f(y), x - y\rangle$. Thus

$$\begin{aligned} \frac{g(y+tw) - g(y)}{t} \geq &-\frac{f(y+tw) - f(y)}{t} - \frac{\langle \nabla f(y+tw) - \nabla f(y), x - y\rangle}{t} \\ &+ \langle \nabla f(y+tw), w\rangle. \end{aligned}$$

Taking $\varliminf_{t \downarrow 0}$, we obtain $\mathrm{d}g(y)(w) \geq -\langle \nabla^2 f(y) w, x - y\rangle = \langle \nabla^2 f(y)(y - x), w\rangle$ and this implies

$$\mathrm{d}g(y)(w) \geq \max\langle \nabla^2 f(y)(y - \overleftarrow{Q}_C(y)), w\rangle. \tag{19}$$

Now take $x_t \in \overleftarrow{Q}_C(y+tw)$ and estimate $g(y+tw) = f(x_t) - f(y+tw) - \langle \nabla f(y+tw), x_t - (y+tw)\rangle$ and $g(y) \geq f(x_t) - f(y) - \langle \nabla f(y), x_t - y\rangle$. Thus

$$\begin{aligned} \frac{g(y+tw) - g(y)}{t} \leq &-\frac{f(y+tw) - f(y)}{t} - \frac{\langle \nabla f(y+tw) - \nabla f(y), x_t - y\rangle}{t} \\ &+ \langle \nabla f(y+tw), w\rangle. \end{aligned} \tag{20}$$

Proposition 4.1(ii) implies that as $t \downarrow 0$, all cluster points of $(x_t)_{t>0}$ lie in $\overleftarrow{Q}_C(y)$. Take a positive sequence $(t_n)_{n \in \mathbb{N}}$ such that $t_n \downarrow 0$ and

$$\mathrm{d}g(y)(w) = \lim_{n \to \infty} \frac{g(y + t_n w) - g(y)}{t_n}.$$

After taking a subsequence if necessary, we also assume that $x_{t_n} \to x \in \overleftarrow{Q}_C(y)$. Then (20) implies that for every $n \in \mathbb{N}$,

$$\frac{g(y+t_n w)-g(y)}{t_n} \leq -\frac{f(y+t_n w)-f(y)}{t_n} - \frac{\langle \nabla f(y+t_n w)-\nabla f(y), x_{t_n}-y\rangle}{t_n} + \langle \nabla f(y+t_n w), w\rangle.$$

Taking limits, we deduce that

$$\begin{aligned} \mathrm{d}g(y)(w) &\leq -\langle \nabla f(y), w\rangle - \langle \nabla^2 f(y) w, x-y\rangle + \langle \nabla f(y), w\rangle \\ &= \langle \nabla^2 f(y)(y-x), w\rangle \leq \max\langle \nabla^2 f(y)(y-\overleftarrow{Q}_C(y)), w\rangle. \end{aligned} \tag{21}$$

Combining (19) and (21), we obtain

$$(\forall w \in \mathbb{R}^J) \quad \mathrm{d}g(y)(w) = \max\langle \nabla^2 f(y)(y-\overleftarrow{Q}_C(y)), w\rangle,$$

from which

$$\hat{\partial} g(y) = \nabla^2 f(y)(y - \operatorname{conv} \overleftarrow{Q}_C(y)).$$

Since $\overleftarrow{Q}_C : U \rightrightarrows C$ is upper semicontinuous and compact-valued by Proposition 4.1(ii), we see that $\operatorname{conv} \overleftarrow{Q}_C : U \rightrightarrows \operatorname{conv} C$ is also upper semicontinuous (see, e.g., [19, Lemma 7.12]). Invoking now the continuity of $\nabla^2 f$, it follows that $\partial_L g(y) = \overline{\lim}_{z\to y} \hat{\partial} g(z) = \nabla^2 f(y)[y - \operatorname{conv} \overleftarrow{Q}_C(y)]$. Proposition 3.2 shows that g is locally Lipschitz on U. Using [22, Theorem 8.49], we deduce that

$$\overline{\partial} g(y) = \operatorname{conv} \partial_L g(y) = \partial_L g(y) = \nabla^2 f(y)[y - \operatorname{conv} \overleftarrow{Q}_C(y)]$$

and

$$(\forall w \in \mathbb{R}^J) \quad \hat{\mathrm{d}} g(y)(w) = \max\langle \nabla^2 f(y)[y - \operatorname{conv} \overleftarrow{Q}_C(y)], w\rangle,$$

which completes the proof. ■

Corollary 5.2. *Suppose that f is twice continuously differentiable on U and that for every $y \in U$, $\nabla^2 f(y)$ is positive definite. Let $y \in U$. Then the following hold.*

(i) *The function $\overleftarrow{F}_C$ is differentiable at $y \in U$ if and only if $\overleftarrow{Q}_C(y)$ is a singleton.*
(ii) *The set $\{y \in U \mid \overleftarrow{Q}_C(y)$ is a singleton$\}$ is residual in U, and its complement in U has Lebesgue measure zero.*

Proof. (i) Assume first that $\overleftarrow{F}_C$ is differentiable at $y \in U$. Then $\hat{\partial} \overleftarrow{F}_C(y) = \{\nabla \overleftarrow{F}_C(y)\}$, and Proposition 5.1 yields

$$\nabla \overleftarrow{F}_C(y) = \nabla^2 f(y)[y - \operatorname{conv} \overleftarrow{Q}_C(y)].$$

Since $\nabla^2 f(y)$ is invertible,

$$\operatorname{conv} \overleftarrow{Q}_C(y) = y - \nabla^2 f(y)^{-1} \nabla \overleftarrow{F}_C(y);$$

thus, $\overleftarrow{Q}_C(y)$ must be a singleton. Conversely, assume that $\overleftarrow{Q}_C(y)$ is a singleton. Apply Proposition 5.1 to deduce that the limiting subdifferential $\partial_L \overleftarrow{F}_C(y)$ is a singleton. This implies that $\overleftarrow{F}_C$ is strictly differentiable at y (see [22, Theorem 9.18(b)]) and hence differentiable at y.

(ii) Since $\overleftarrow{F}_C$ is locally Lipschitz on U (see Proposition 3.2), Rademacher's Theorem (see [4, Theorem 9.1.2] or [9, Corollary 3.4.19]) guarantees that $\overleftarrow{F}_C$ is differentiable almost everywhere on U. Moreover, since $\overleftarrow{F}_C$ is Clarke regular on U (see Proposition 5.1), we use [15, Theorem 10] to deduce that $\overleftarrow{F}_C$ is generically differentiable on U. The result now follows from (i). ■

6. Characterizations

In this section, we give complete characterizations of sets with unique farthest-point properties. To do so, we need the following two key results on the expression of the convex-analytical subdifferential of the function $-f^\vee + \iota_{-C}$ (see also (12)) and of the conjugate $(-f^\vee + \iota_{-C})^*$ in terms of $\overleftarrow{Q}_C \circ \nabla f^*$. These results extend Hiriart-Urruty's [11, Proposition 4.4 and Corollary 4.5] to the framework of Bregman distances.

Lemma 6.1. *Let $x \in -C$. Then $\partial(-f^\vee + \iota_{-C})(x) = (\overleftarrow{Q}_C \circ \nabla f^*)^{-1}(-x)$.*

Proof. Let $s \in \mathbb{R}^J$. By [13, Theorem X.1.4.1], $s \in \partial(-f^\vee + \iota_{-C})(x)$ if and only if

$$-f(-x) + \iota_{-C}(x) + (-f^\vee + \iota_{-C})^*(s) = \langle s, x\rangle. \tag{22}$$

In view of (14), equation (22) is equivalent to $-f(-x) + (\overleftarrow{F}_C \circ \nabla f^*)(s) - f^*(s) = \langle x, s\rangle$, and hence to $\overleftarrow{F}_C(\nabla f^*(s)) = f(-x)+f^*(s)+\langle x, s\rangle = D(-x, \nabla f^*(s))$, i.e., to $-x \in \overleftarrow{Q}_C(\nabla f^*(s))$. ■

Lemma 6.2. *We have $\partial(-f^\vee + \iota_{-C})^* = -\operatorname{conv}(\overleftarrow{Q}_C \circ \nabla f^*)$.*

Proof. Let x and s be in $\mathbb{R}^J$. By [13, Lemma X.1.5.3] or [22, Corollary 3.47],

$$\overline{\operatorname{conv}}(-f^\vee + \iota_{-C}) = \operatorname{conv}(-f^\vee + \iota_{-C}).$$

On the other hand,

$$x \in \partial(-f^\vee + \iota_{-C})^*(s) \Leftrightarrow s \in \partial(-f^\vee + \iota_{-C})^{**}(x) = \partial\, \overline{\operatorname{conv}}(-f^\vee + \iota_{-C})(x).$$

Altogether,

$$x \in \partial(-f^\vee + \iota_{-C})^*(s) \Leftrightarrow s \in \partial \operatorname{conv}(-f^\vee + \iota_{-C})(x). \tag{23}$$

Now by [13, Theorem X.1.5.6], $s \in \partial \operatorname{conv}(-f^\vee + \iota_{-C})(x)$ if and only if there there exist nonnegative real numbers $\lambda_1, \ldots, \lambda_{J+1}$ and points $x_1, \ldots, x_{J+1}$ in $\mathbb{R}^J$ such that

$$\sum_{j=1}^{J+1} \lambda_j = 1, \quad x = \sum_{j=1}^{J+1} \lambda_j x_j \quad \text{and} \quad s \in \bigcap_{j:\, \lambda_j > 0} \partial(-f^\vee + \iota_{-C})(x_j);$$

furthermore, Lemma 6.1 shows that $s \in \partial(-f^\vee + \iota_{-C})(x_j) \Leftrightarrow x_j \in -(\overleftarrow{Q}_C \circ \nabla f^*)(s)$. Therefore, the two conditions of (23) are also equivalent to $x \in -\sum_{j=1}^{J+1} \lambda_j (\overleftarrow{Q}_C \circ \nabla f^*)(s)$. ■

Remark 6.3. When $f = \frac{1}{2}\|\cdot\|^2$ is the (halved) energy (see Example 2.1(i)), then (14) turns into

$$(-f^\vee + \iota_{-C})^* = \frac{1}{2}\Delta_C^2 - \frac{1}{2}\|\cdot\|^2,$$

where $\Delta_C \colon x \mapsto \sup \|x - C\|$. In this case, the conclusion of Lemma 6.2 is classic; see [12, pages 262–264] and [11, Theorem 4.3].

We need the following result from [24] (see also [26, Section 3.9]).

Fact 6.4 (*Soloviov*). Let $g : \mathbb{R}^J \to \,]-\infty, +\infty]$ be lower semicontinuous and such that g^* is essentially smooth. Then g is convex.

We are now ready for the main result of this section.

Theorem 6.5 (*Characterizations of $\overleftarrow{D}$-Klee Sets*). *The following are equivalent.*

(i) *C is $\overleftarrow{D}$-Klee, i.e., $\overleftarrow{Q}_C$ is a single-valued on U.*
(ii) *$\overleftarrow{Q}_C$ is single-valued and continuous on U.*
(iii) *$\overleftarrow{F}_C \circ \nabla f^*$ is continuously differentiable on $\mathbb{R}^J$.*
(iv) *$-f^\vee + \iota_{-C}$ is convex.*
(v) *C is a singleton.*

If (i)–(v) *hold, then*

$$\nabla(\overleftarrow{F}_C \circ \nabla f^*) = \nabla f^* - \overleftarrow{Q}_C \circ \nabla f^*. \tag{24}$$

Moreover, if f is twice continuously differentiable and the Hessian $\nabla^2 f(y)$ is positive definite for every $y \in U$, then (i)–(v) *are also equivalent to*

(vi) *$\overleftarrow{F}_C$ is differentiable on U,*

in which case $\overleftarrow{F}_C$ is actually continuously differentiable on U with

$$(\forall y \in U) \quad \nabla \overleftarrow{F}_C(y) = \nabla^2 f(y) \left(y - \overleftarrow{Q}_C(y)\right). \tag{25}$$

Proof. "(i) $\Rightarrow$ (ii): Apply Proposition 4.1(ii). "(ii) $\Rightarrow$ (iii)": On the one hand, (14) implies

$$\overleftarrow{F}_C \circ \nabla f^* = (-f^\vee + \iota_{-C})^* + f^*. \tag{26}$$

On the other hand, Lemma 6.2 yields

$$\nabla(-f^\vee + \iota_{-C})^* = -\overleftarrow{Q}_C \circ \nabla f^*. \tag{27}$$

Combining (26) and (27), we obtain altogether (iii), and also (24). "(iii)$\Rightarrow$(iv)": This follows from (14) and Fact 6.4. "(iv)$\Rightarrow$(v)": Assume to the contrary that C is not a singleton, fix two distinct points y_0 and y_1 in C, and $t \in \mathbb{R}$ with $0 < t < 1$. Set $y_t := (1-t)y_0 + ty_1$. Since $-f^\vee + \iota_{-C}$ is a convex function, its domain $-C$ is a convex set. Hence $y_t \in C$ and $-f(y_t) \le -(1-t)f(y_0) - tf(y_1)$. However, since f is strictly convex, the last inequality is impossible. Therefore, C is a singleton. "(v)$\Rightarrow$(i)": This is obvious.

Finally, we assume that f is twice differentiable on U and that the $\nabla^2 f(y)$ is invertible, for every $y \in U$. The equivalence of (i) and (vi) follows from Corollary 5.2(i), and (18) yields the formula for gradient (25), which is continuous by (ii). ■

Theorem 6.6. *Set*

$$\theta_C : \mathbb{R}^J \to \,]-\infty, +\infty] : x \mapsto \inf_{c \in C} \left(f(x+c) - f(c) \right). \tag{28}$$

Then θ_C is proper and lower semicontinuous,

$$\theta_C = f \,\Box\, \left(-f^\vee + \iota_{-C}\right), \tag{29}$$

where this infimal convolution is exact at every point in $\operatorname{dom}\theta_C = \operatorname{dom} f - C$*, and*

$$\theta_C^* = \overleftarrow{F}_C \circ \nabla f^*. \tag{30}$$

Moreover,

$$\theta_C \text{ is convex} \quad \Leftrightarrow \quad C \text{ is a singleton.} \tag{31}$$

Proof. For every $x \in \mathbb{R}^J$, we have

$$\begin{aligned}\big(f\Box(-f^\vee + \iota_{-C})\big)(x) &= \inf_y (f(x-y) - f(-y) + \iota_{-C}(y)) \\ &= \inf_{-y\in C} (f(x-y) - f(-y)) \\ &= \inf_{c\in C} (f(x+c) - f(c)) \\ &= \theta_C(x),\end{aligned}$$

which verifies (29) and the domain formula. Since $\operatorname{dom}(-f^\vee + \iota_{-C}) = -C$ is bounded, [22, Proposition 1.27] implies that $f\Box(-f^\vee + \iota_{-C})$ is proper and lower semicontinuous, and that the infimal convolution is exact at every point in its domain. Using (14) and [22, Theorem 11.23(a)], we obtain

$$\overleftarrow{F}_C \circ \nabla f^* = f^* + (-f^\vee + \iota_{-C})^* = \big(f\Box(-f^\vee + \iota_{-C})\big)^*. \tag{32}$$

This and (29) yield (30).

It remains to prove (31). The implication "⇐" is clear. We now tackle "⇒". Since $U - C \subseteq \operatorname{dom} f - C = \operatorname{dom}\theta_C$ and since $C \subset U$, we have $0 \in \operatorname{int}\operatorname{dom}\theta_C$. Take $x \in \operatorname{dom}\partial\theta_C$ and $x^* \in \partial\theta_C(x)$. Then

$$(\forall y \in \mathbb{R}^J) \quad \langle x^*, y - x\rangle \le \theta_C(y) - \theta_C(x). \tag{33}$$

On the other hand, there exists $\bar{c} \in C$ such that $\theta_C(x) = f(x+\bar{c}) - f(\bar{c})$ and also $(\forall y \in \mathbb{R}^J)\theta_C(y) \le f(y+\bar{c}) - f(\bar{c})$. Altogether,

$$(\forall y \in \mathbb{R}^J) \quad \langle x^*, y - x\rangle \le f(y+\bar{c}) - f(x+\bar{c}), \tag{34}$$

and this implies $x^* \in (\partial f(\cdot + \bar{c}))(x)$. Since f is essentially smooth, it follows that $\partial\theta_C(x)$ is a singleton. In view of [21, Theorem 26.1], θ_C is essentially smooth, and thus differentiable on $\operatorname{int}\operatorname{dom}\theta_C$. Since $0 \in \operatorname{int}\operatorname{dom}\theta_C$, θ_C is locally Lipschitz and differentiable at every point in an open neighbourhood V of 0. Now set

$$g\colon \mathbb{R}^J \to [-\infty, +\infty[\ : x \mapsto \sup_{c\in C} (f(c) - f(c+x)). \tag{35}$$

Then $\theta_C = -g$, g is lower C^1 (see [22, Definition 10.29]), and

$$C = \{c \in C \mid g(0) = 0 = f(c) - f(c)\}.$$

By [22, Theorem 10.31], $\overline{\partial}g(0) = \operatorname{conv}\{-\nabla f(c) \mid c \in C\} = -\operatorname{conv}\{\nabla f(C)\}$. As g is locally Lipschitz on V, [8, Theorem 2.3.1] now yields

$$\overline{\partial}(-g)(0) = -\overline{\partial}g(0) = \operatorname{conv}\{\nabla f(C)\}.$$

Using finally that $\theta_C = -g$ is convex, and that $\partial = \overline{\partial}$ for convex functions, [8, Proposition 2.2.7], we obtain

$$\nabla\theta_C(0) = \partial\theta_C(0) = \partial(-g)(0) = \overline{\partial}(-g)(0) = \operatorname{conv}\{\nabla f(C)\},$$

i.e., $\operatorname{conv}\{\nabla f(C)\} = \nabla\theta_C(0)$. Therefore, $\nabla f(C)$ is a singleton, and so is C by Fact 2.2. ■

Remark 6.7. If $f = \frac{1}{2}\|\cdot\|^2$, then

$$\begin{aligned}\theta_C(x) &= \inf_{c\in C}\left(\frac{1}{2}\|x+c\|^2 - \frac{1}{2}\|c\|^2\right)\\ &= \inf_{c\in C}\left(\frac{1}{2}\|x\|^2 + \langle x, c\rangle\right)\\ &= \frac{1}{2}\|x\|^2 - \sup\langle -C, x\rangle\end{aligned}$$

is the function introduced by Hiriart-Urruty in [11, Definition 4.1]. Thus, equivalence (31) extends [11, Proposition 4.2].

7. Right-Bregman-farthest-point maps

In this section, we relax our assumptions on f, i.e., we will only assume **A1** and **A3**. It will be important to emphasis the dependence on f for the Bregman distance and for the (left and right) Bregman-farthest-point map; consequently, we will write D_f, $\overleftarrow{Q}_C^f$, $\overrightarrow{Q}_C^f$, and similarly for f^*. While D_f is generally not convex in its right (second) argument – which makes the theory asymmetric – it turns out that $\overrightarrow{Q}_C^f$ can be studied via $\overleftarrow{Q}_{\nabla f(C)}^{f^*}$.

Proposition 7.1. *Suppose that f and C satisfy* **A1** *and* **A3**. *Then*

$$\overrightarrow{Q}_C^f = \nabla f^* \circ \overleftarrow{Q}_{\nabla f(C)}^{f^*} \circ \nabla f \quad \text{and} \quad \overleftarrow{Q}_{\nabla f(C)}^{f^*} = \nabla f \circ \overrightarrow{Q}_C^f \circ \nabla f^*. \tag{36}$$

Proof. Applying [2, Theorem 3.7(v)] to f^*, we see that

$$(\forall x^* \operatorname{int}\operatorname{dom} f^*)(\forall y^* \in \operatorname{int}\operatorname{dom} f^*) \quad D_{f^*}(x^*, y^*) = D_f\left(\nabla f^*(y^*), \nabla f^*(x^*)\right).$$

Hence for every $y^* \in \operatorname{int}\operatorname{dom} f^*$, we obtain

$$\begin{aligned}\overleftarrow{Q}_{\nabla f(C)}^{f^*}(y^*) &= \operatorname*{argmax}_{x^*\in\nabla f(C)} D_{f^*}(x^*, y^*)\\ &= \operatorname*{argmax}_{x^*\in\nabla f(C)} D_f\left(\nabla f^*(y^*), \nabla f^*(x^*)\right)\\ &= \nabla f\left(\overrightarrow{Q}_{\nabla f^*(\nabla f(C))}^f\left(\nabla f^*(y^*)\right)\right)\\ &= \nabla f\left(\overrightarrow{Q}_C^f\left(\nabla f^*(y^*)\right)\right),\end{aligned}$$

and this is the right identity in (36); the left one now follows Fact 2.2. ■

Theorem 7.2. *Suppose that f and C satisfy* **A1** *and* **A3**, *that* $\operatorname{dom} f = \mathbb{R}^J$, *and that C is $\overrightarrow{D}$-Klee, i.e., for every $y \in \mathbb{R}^J$, $\overrightarrow{Q}_C^f(y)$ is a singleton. Then C is a singleton.*

Proof. Since C is compact and $\nabla f : \mathbb{R}^J \to \operatorname{int}\operatorname{dom} f^*$ is an isomorphism (see Fact 2.2), we deduce that $\nabla f(C)$ is a compact subset of $\operatorname{int}\operatorname{dom} f^*$. Furthermore, by (36), $\nabla f(C)$ is $\overleftarrow{D}$-Klee with respect to f^*. Since f^* satisfies **A1–A3**, we apply Theorem 6.5 and conclude that $\nabla f(C)$ is a singleton. Finally, again using Fact 2.2, we see that C is a singleton. ■

Remark 7.3. We do not know whether Theorem 7.2 is true if the full-domain assumption on f is dropped.

Acknowledgments

Heinz Bauschke was partially supported by the Natural Sciences and Engineering Research Council of Canada and by the Canada Research Chair Program. Xianfu Wang was partially supported by the Natural Sciences and Engineering Research Council of Canada. Jane Ye was partially supported by the Natural Sciences and Engineering Research Council of Canada. Xiaoming Yuan was partially supported by the Pacific Institute for the Mathematical Sciences, the University of Victoria, the University of British Columbia Okanagan, and the National Science Foundation of China Grant 10701055.

References

[1] E. Asplund, Sets with unique farthest points, Israel J. Math. 5 (1967) 201–209.
[2] H.H. Bauschke, J.M. Borwein, Legendre functions and the method of random Bregman projections, J. Convex Anal. 4 (1997) 27–67.
[3] H.H. Bauschke, X. Wang, J. Ye, X. Yuan, Bregman distances and Chebyshev sets, J. Approx. Theory. Preprint http://arxiv.org/abs/0712.4030v1, December 24, 2007, in press (doi:10.1016/j.jat.2008.08.014).
[4] J.M. Borwein, A.S. Lewis, Convex Analysis and Nonlinear Optimization, 2nd edition, Springer, New York, 2006.
[5] L.M. Bregman, The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming, U.S.S.R. Comp. Math. Math. Phys. 7 (1967) 200–217.
[6] D. Butnariu, A.N. Iusem, Totally Convex Functions for Fixed Point Computation in Infinite Dimensional Optimization, Kluwer, Dordrecht, 2000.
[7] Y. Censor, S.A. Zenios, Parallel Optimization, Oxford University Press, 1997.
[8] F.H. Clarke, Optimization and Nonsmooth Analysis, Wiley Interscience, New York, 1983.
[9] F.H. Clarke, Yu.S. Ledyaev, R.J. Stern, P.R. Wolenski, Nonsmooth Analysis and Control Theory, Springer-Verlag, New York, 1998.
[10] F. Deutsch, Best Approximation in Inner Product Spaces, Springer-Verlag, New York, 2001.
[11] J.-B. Hiriart-Urruty, La conjecture des points les plus éloignés revisitée, Ann. Sci. Math. Québec 29 (2005) 197–214.
[12] J.-B. Hiriart-Urruty, Potpourri of conjectures and open questions in nonlinear analysis and optimization, SIAM Rev. 49 (2007) 255–273.
[13] J.-B. Hiriart-Urruty, C. Lemaréchal, Convex Analysis and Minimization Algorithms II, Springer, New York, 1996.
[14] V. Klee, Convexity of Chevyshev sets, Math. Ann. 142 (1960/61) 292–304.
[15] P.D. Loewen, X. Wang, On the multiplicity of Dini subgradients in separable spaces, Nonlinear Anal. 58 (2004) 1–10.
[16] B.S. Mordukhovich, Variational Analysis and Generalized Differentiation I, Springer-Verlag, Berlin, 2006.
[17] T.S. Motzkin, E.G. Straus, F.A. Valentine, The number of farthest points, Pacific J. Math. 3 (1953) 221–232.
[18] B.B. Panda, O.P. Kapoor, On farthest points of sets, J. Math. Anal. Appl. 62 (1978) 345–353.
[19] R.R. Phelps, Convex Functions Monotone Operators and Differentiability, 2nd edition, in: Lecture Notes in Mathematics, vol. 1364, Springer-Verlag, 1993.
[20] F.P. Preparata, M.I. Shamos, Computational Geometry, Springer-Verlag, 1993.
[21] R.T. Rockafellar, Convex Analysis, Princeton University Press, Princeton, 1970.
[22] R.T. Rockafellar, R.J.-B. Wets, Variational Analysis, Springer-Verlag, New York, 1998.
[23] S. Simons, Minimax and Monotonicity, in: Lecture Notes in Mathematics, vol. 1693, Springer-Verlag, 1998.
[24] V. Soloviov, Duality for nonconvex optimization and its applications, Anal. Math. 19 (1993) 297–315.
[25] U. Westphal, T. Schwartz, Farthest points and monotone operators, Bull. Austral. Math. Soc. 58 (1998) 75–92.
[26] C. Zălinescu, Convex Analysis in General Vector Spaces, World Scientific Publishing, 2002.